

# Perturbative solution to the SIS epidemic on networks

Lloyd P. Sanders,<sup>\*</sup> Dirk Brockmann,<sup>†</sup> and Tobias Ambjörnsson<sup>‡</sup>

We provide a closed form perturbative solution to the  $M$ -city network SIS model using the transport rates between cities as a perturbation parameter. We calculate explicitly the first-order perturbation, and quantify that validity of the approximation with respect to the number of initially infected and the mobility rates between cities.

To further corroborate the results, we simulate and analyze a highly infectious SIS outbreak in New Zealand (NZ, *Aotearoa*), where we illustrate how the first-order expansion compares to the numerical solution, when we recompose NZ into two subpopulations: Auckland (*Tāmaki Makaurau*, containing over one third of the total population), and the remainder of NZ. In this comparison we use documented air traffic data, and isolate the country from international travel. We demonstrate that within the first-order solution's range, it compares well to that of the numerical simulations.

## INTRODUCTION

In mathematical epidemiology the susceptible-infected-susceptible (SIS) model is the one of the most elementary compartmental models, receiving consistent attention (the specifics of which are given at length below) since the seminal work of Kermack and McKendrick [1]. Put simply, this model partitions a large, homogeneous population into two compartments: susceptible and infected. The population is a well mixed system, and the birth and death rate are neglected. When a susceptible person comes into contact with an infected person, there is some finite probability that the susceptible person may become infected. An infected person will lose their infection after some typical time, returning back to the susceptible health state.

The simplicity of this model and its ability to characterize the main motifs of viral infections where recovery does not assure immunity (for example Gonorrhea [2] or Chlamydia [3]), has allowed for extensive research. Due to the mathematical tractability of the model, many extensions have been applied, to include other important dynamical factors; recent investigations include: analyzing the integrability of the SIS model with so-called vital dynamics [4] (explicit inclusion of birth and death rates, where new-borns are usually birthed into the susceptible compartment); understanding time-delayed infection periods [5]; and how intrinsic noise on the transmission rate may affect the dynamics [6]. Although these models have focused on deterministic mean-field approaches as we shall herein, there is also a surge in converting the models over to their stochastic counter parts [7], and analyzing different properties of the system, for a recent example see [8].

Over the last two decades the literature has seen a profusion of graph/network theory research (for a review, see [9]), from which there has been intense focus on how this vogue field may contribute to the understanding the large scale realism of human mobility [10, 11], and in turn comprehending the etiology of epidemics on these systems [12]. Within approximately the same period

people have incorporated the concept of metapopulations [13] (a population of populations where in each, mean-field equations suffice to describe the system dynamics) into theoretical epidemiology [14]; where interesting recent examples include the work by Lund et. al. [15] in investigating the effects of heterogeneity on epidemic spreading in metapopulations; or the work by Colizza and Vespignani on the invasion threshold in heterogeneous metapopulation networks [16].

These similar directions of implicit/explicit spatial structure incorporation have naturally been applied to SIS models. Of late, network theory has been applied to the dynamics of an SIS virus on random networks [17], how non-Markovian infection changes these dynamics on a network [18], or simply exploring the full parameter space of the SIS dynamics on a network via large scale numerical simulations [19]. Concerning metapopulations, Arrigoni and Pugliese considered a stochastic  $M$ -city,  $N$ -inhabitant SIS model [20] and Ball analyzed the stochastic and deterministic SIS model within a population of households [21]. Essentially the current state of the research has implied two salient points: computational power is easily accessible, and the network epidemic modeling is not readily amenable to classical mathematical tools. In this manuscript we hope to address these topics, whereby we amalgamate the canonical deterministic SIS model with the current impetus toward network/metapopulation modeling through the use of perturbation theory, to quantify the effects of human mobility on an arbitrary network through analytical calculations, substantiated by the equivalent numerical simulations.

Within the following section we review the analytics of the canonical single-city SIS model, after which we segregate the population into an  $M$ -city network(/metapopulation), whereupon the SIS infection is introduced separately to each. We present a closed-form recursive perturbative solution to the network model, therefrom we calculate explicitly the first-order perturbation leading to our study's main result Eq. (20). Subsequently we compare our result to a test case scenario using real-world population and air traffic data. We then

discuss the benefits and limitations of the model and where this work may be applied and built upon.

## SIS MODEL

In this section we will describe the equations which govern the single- and  $M$ -city models and perform analytical analysis where applicable. Then, the results are compared to numerical simulations.

### Canonical single-city SIS model

The single-city SIS model considers a large, well mixed, population,  $N$ , in some closed environment where the mortality and birth rates are neglected. The populace is divided into two compartments: susceptible,  $S$ ; and infected,  $I$ ; where  $N = S + I = \text{constant}$ . Susceptibles may become infected from contact with the infected at a rate  $\beta$ , and the infected compartment of the population will lose constituents at a rate  $\gamma$ . The mean-field dynamics of each state is then described by the set of equations:

$$\partial_t S(t) = -\frac{\beta}{N}SI + \gamma I, \quad (1)$$

$$\partial_t I(t) = \frac{\beta}{N}SI - \gamma I, \quad (2)$$

where initial conditions are:  $S(t = 0) = S_0 > 0$ , and  $I(t = 0) = I_0 > 0$ . One immediately notes that  $\partial_t [S(t) + I(t)] = 0$ , which then ensures that the total population,  $N$ , is constant for all time.

Solving Eqs. (1) and (2) analytically, the solution - a logistic growth function - is well known (for example see [2]), but is included here for completeness. As this is a closed system, it is sufficient to consider only the solution to the infected population - since  $S(t) = N - I(t)$ . Therefore, we can recast Eq. (2) into:  $\partial_t I(t) = \beta(N - I)I/N - \gamma I$ , and make a transformation:  $y = I^{-1}$  (such that  $I \neq 0$ ), where  $dy/dI = -I^{-2}$ . Therefore Eq. (2) can be written as  $\partial_t y(t) = -\chi y + \beta/N$ , where  $\chi = \beta - \gamma$ . The solution is thus  $y(t) = Ce^{-\chi t} + \beta/(N\chi)$ , where  $C = y_0 - \beta/(\chi N)$ , and  $y_0 = 1/I_0$ . Reversing the transformation, we find the solution to  $I(t)$ ,

$$I(t) = \frac{I_\infty}{1 + Ve^{-\chi t}}, \quad (3)$$

where  $I_\infty = \chi N/\beta$ , is the stable/endemic state of the infected population; and  $V = I_\infty/I_0 - 1$ . With respect to Eq. (3), for an epidemic to take place (i.e. some finite fraction of the population remains infected in the long-time limit), we require the basic reproductive ratio:  $R_0 = \beta/\gamma$  to satisfy:  $R_0 > 1$  (which will be assumed henceforth).

To incorporate a spatial component to the model let us consider an arbitrary network of cities.

### Perturbative solution to the $M$ -city SIS model

We here further the result found in the previous section through incorporation of an implicit spatial component, by stratifying the large population into  $M$  subpopulations (caricatured by Fig. 1). Within each subpopulation, which may be regarded as a city (but just as readily considered to be a community, town, or country), the same assumptions stated in the canonical model still hold: the population is sufficiently large, and well mixed. One then allows for mobility between the nodes (cities) on the network, where the fraction of persons traveling from city  $j$  to  $i$  per unit time is given by the transport rate,  $\omega_{i \leftarrow j}$ , where detailed balance [12] is assumed:

$$w_{j \leftarrow i} N_i = w_{i \leftarrow j} N_j, \quad (4)$$

guaranteeing the total population  $N_i$  of city  $i$  is constant for all time.

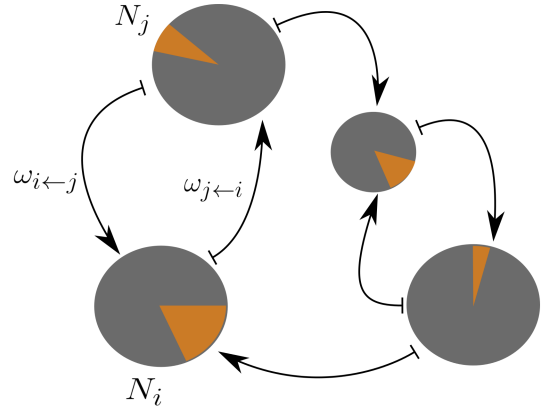


FIG. 1. Caricature of an arbitrary network of communities, infected with an SIS type virus, described by Eqs. (5) and (6), where gray denotes susceptible people, orange, the infected and the arrows show possible mobility between nodes.  $N_i$  is the total population of city  $i$ , and  $\omega_{j \leftarrow i}$  is the travel rate, the fraction of the population that travels from city  $i$  to  $j$  per unit time.

Upon each city, an SIS virus is introduced, where the  $i^{\text{th}}$  city has an infectivity rate of  $\beta_i$ , and recovery rate  $\gamma_i$ . The mean-field set of equations that then describe the dynamics are given by

$$\partial_t S_i(t) = -\frac{\beta_i}{N_i} S_i I_i + \gamma_i I_i + \varepsilon \sum_{j=1}^M (w_{i \leftarrow j} S_j - w_{j \leftarrow i} S_i), \quad (5)$$

$$\partial_t I_i(t) = \frac{\beta_i}{N_i} S_i I_i - \gamma_i I_i + \varepsilon \sum_{j=1}^M (w_{i \leftarrow j} I_j - w_{j \leftarrow i} I_i). \quad (6)$$

where we have defined  $\omega_{i \leftarrow i} = 0$ , and  $\varepsilon (= 1)$  is a convenient parameter introduced here for counting perturbation orders (see below). Summing Eqs. (5) and (6), we find:  $\partial_t N_i = 0$ , as required by detailed balance.

To begin the derivation of the perturbative solution, we first assume the influx and outflux of citizens from a given city is small (defined quantitatively later), from there we can define a perturbative solution in terms of the travel rates, namely

$$I_i(t) = \sum_{k=0}^{\infty} \varepsilon^k I_i^{(k)}(t) = I_i^{(0)}(t) + \sum_{k=1}^{\infty} \varepsilon^k I_i^{(k)}(t), \quad (7)$$

where  $I_i^{(0)}(t)$  is given by Eq. (3), with replacements  $\beta \rightarrow \beta_i$  and  $\gamma \rightarrow \gamma_i$  (and therefore  $\chi \rightarrow \chi_i$ ). Explicitly:

$$I_i^{(0)}(t) = \frac{I_{\infty,i}}{1 + V_i e^{-\chi_i t}}. \quad (8)$$

Similarly we can define the perturbative solution to the number of susceptibles in city  $i$  as  $S_i = \sum_{k=0}^M \varepsilon^k S_i^{(k)}$ , which then implies, due to detailed balance, that

$$I_i^{(k)} = -S_i^{(k)} \quad \text{for } k \geq 1. \quad (9)$$

Substituting Eqs. (9) and (7) into Eq. (6) we find

$$\begin{aligned} \partial_t \sum_k \varepsilon^k I_i^{(k)} &= \frac{\beta_i}{N_i} \left( \sum_{k,l} \varepsilon^k \varepsilon^l [-I_i^{(k)}] I_i^{(l)} \right) \\ &\quad - \gamma_i \sum_k \varepsilon^k I_i^{(k)} + \sum_j \left( w_{i \leftarrow j} \sum_k \varepsilon^{k+1} I_j^{(k)} \right. \\ &\quad \left. - w_{j \leftarrow i} \sum_k \varepsilon^{k+1} I_i^{(k)} \right) \end{aligned} \quad (10)$$

Equating factors of  $\varepsilon^k$ , we find for  $k = 0$ , that

$$\partial_t I_i^{(0)} = \chi_i I_i^{(0)} - \frac{\beta_i}{N_i} [I_i^{(0)}]^2, \quad (11)$$

which is equivalent to Eq. (2), and whose solution is therefore given by Eq. (8), with appropriate  $i$ -dependencies of the virus parameters. For  $k \geq 1$  we obtain our formal perturbation equations

$$\begin{aligned} \partial_t I_i^{(k)} - \left( \chi_i - \frac{2\beta_i I_i^{(0)}}{N_i} \right) I_i^{(k)} &= -\frac{\beta_i}{N_i} \left( \sum_{k'=1}^{k-1} I_i^{(k-k')} \right) \\ &\quad + \sum_j \left( w_{i \leftarrow j} I_j^{(k-1)} - w_{j \leftarrow i} I_i^{(k-1)} \right). \end{aligned} \quad (12)$$

Thus, we have formally converted the non-linear problem in Eqs. (5) and (6), into a set of inhomogeneous, linear equations - Eq. (12) - with time dependent coefficients. The time dependence of these coefficients enters only through the known quantity,  $I_i^{(0)}(t)$ , whereas the right-hand side depends recursively on the previous perturbation orders.

We proceed by expressing a formal solution to Eq. (12) though employment of the integration factor method. Firstly, we define the so-called integration factor:  $\exp(B_i(t))$ , where  $B_i(t) =$

$-\int_0^t [\chi_i - 2\beta_i I_i^{(0)}(t')/N_i] dt'$ . Then the formal solution to the  $k^{\text{th}}$ -order perturbative term is  $I_i^{(k)}(t) = \exp(-B_i(t)) \left[ \int_0^t \exp(B_i(t')) g_i(t') dt' + G \right]$ , where from the initial conditions:  $I_i^{(k)}(t=0) = 0$ , we have  $G = 0$ . The function  $g_i(t)$  is defined as

$$\begin{aligned} g_i^{(k-1)}(t) &= -\frac{\beta_i}{N_i} \left( \sum_{k'=1}^{k-1} I_i^{(k-k')} \right) \\ &\quad + \sum_j \left( w_{i \leftarrow j} I_j^{(k-1)} - w_{j \leftarrow i} I_i^{(k-1)} \right). \end{aligned} \quad (13)$$

Interestingly, we are able to calculate  $B_i(t)$  explicitly. Using Eq. (8), we can write  $B_i(t) = -\chi_i t + (2\beta_i I_{\infty,i})/(N_i) \int_0^\infty (1 + V_i e^{-\chi_i t'})^{-1} dt'$ . We solve this to yield the solution

$$B_i(t) = \ln \left[ e^{\chi_i t} \left( \frac{1 + V_i e^{-\chi_i t}}{1 + V_i} \right)^2 \right]. \quad (14)$$

Using the solution for  $B_i(t)$  and Eq. (13) and substituting this into the formal solution given, we explicitly obtain the  $k^{\text{th}}$  order perturbation, namely

$$\begin{aligned} I_i^{(k)}(t) &= e^{-\chi_i t} (1 + V_i e^{-\chi_i t})^{-2} \times \\ &\quad \left[ \int_0^t e^{\chi_i t'} (1 + V_i e^{-\chi_i t'})^2 g_i^{(k-1)}(t') dt' \right]. \end{aligned} \quad (15)$$

With this closed form expression, we are able to calculate any order perturbation we require, recursively. Namely, starting from the zeroth-order solution, Eq. (8), we can insert this into Eq. (13), the result of which is then input into Eq. (15) to find the first-order perturbation (shown explicitly in the following section). If one is to find the next order, one merely uses the first-order result in place of the zeroth-order solution to, following the outlined algorithm, arrive at the second-order perturbation. This operation may be repeated until the desired number of orders are achieved. Then the orders are summed, viz. Eq. (7), to gain the final solution to the infected population contained in city  $i$ .

### Explicit first-order perturbation

Let us now calculate the first order perturbation term, for which  $k = 1$ . Then function  $g_i^{(0)}$ , see Eq. (13), is explicitly:  $g_i^{(0)}(t) = \sum_j (w_{i \leftarrow j} I_j^{(0)} - w_{j \leftarrow i} I_i^{(0)})$ , such that Eq. (15), using Eq. (8), becomes

$$I_i^{(1)} = \frac{e^{-\chi_i t}}{(1 + V_i e^{-\chi_i t})^2} \sum_j [w_{i \leftarrow j} Q_{ij} - w_{j \leftarrow i} Q_{ii}] \quad (16)$$

where

$$Q_{ij} = I_{\infty,j} \int_0^t e^{\chi_i t'} \frac{(1 + V_i e^{-\chi_i t'})^2}{1 + V_j e^{-\chi_j t'}} dt'. \quad (17)$$

The quantity  $Q_{ij}$  may be expressed in terms of hypergeometric functions [22].

For the scope of this manuscript, let us analyze the case where we shall assume that all virus rate parameters are independent of the city, namely  $\beta_i = \beta_j = \beta$  and  $\gamma_i = \gamma_j = \gamma$ . Explicitly evaluating Eq. (17), we find that

$$Q_{ij} = \frac{-I_{\infty,j}}{\chi} \left[ \frac{(V_i - V_j)^2}{V_j} \ln \left( \frac{1 + V_j e^{-\chi t}}{1 + V_j} \right) - \chi t (2V_i - V_j) + 1 - e^{\chi t} \right]. \quad (18)$$

$$I_i(t) = \frac{I_{\infty,i}}{1 + V_i e^{-\chi t}} \left( 1 + \frac{e^{-\chi t}}{(1 + V_i e^{-\chi t})} \sum_j \frac{\omega_{j \leftarrow i}}{\chi} (V_i - V_j) \left[ \chi t - \frac{V_i - V_j}{V_j} \ln \left( \frac{1 + V_j e^{-\chi t}}{1 + V_j} \right) \right] \right) + \mathcal{O}(w_{i \leftarrow j}^2), \quad (20)$$

where

$$V_i = \frac{I_{\infty,i}}{I_{0,i}} - 1,$$

and

$$I_{\infty,i} = \frac{N_i \chi}{\beta}.$$

The first-order approximation to  $I_i(t)$ , namely. Eq. (20), is valid when the perturbation due to mobility is “small”. To quantify explicitly the validity of Eq. (20) we introduce the validity criterion:

$$C_{j \leftarrow i} = \left| \frac{\omega_{j \leftarrow i}}{\chi} (V_i - V_j) \right| = \left| \frac{\omega_{j \leftarrow i}}{\beta} \left( \frac{N_i}{I_{0,i}} - \frac{N_j}{I_{0,j}} \right) \right|. \quad (21)$$

For a given system, when  $\sum_j C_{j \leftarrow i} \approx 0$  the zeroth-order solution is valid for  $I_i(t)$ ; when  $\sum_j C_{j \leftarrow i} > 1$  the first-order solution breaks down. In particular we point out that, in fact, besides the travel rate  $\omega_{j \leftarrow i}$  (in units of  $\beta$ ), also the fraction of initially infecteds enter Eq. (21) in a non-trivial way [23].

### SIS Epidemic in New Zealand (*Aotearoa*)

We proceed by measuring the performance of our analytical expression, Eq. (20), with realistic population and air transport data through a test case scenario, an SIS epidemic in New Zealand (NZ). Firstly, we suppose a highly infectious virus ( $\beta = 0.25 \text{ day}^{-1}$ ,  $\gamma = 0.025 \text{ day}^{-1}$ ,  $R_0 = 10$ ) has established itself within NZ, and assume the virus is only spread city to city via the air traffic transport network where there is no transport internationally with NZ (which one could think of as a quarantine measure). Then, to understand how the country's

Due to detailed balance, Eq. (4), we know that  $\omega_{i \leftarrow j} I_{\infty,j} = \omega_{j \leftarrow i} I_{\infty,i}$ , using this relation, we can write out the first order perturbation, Eq. (16), explicitly as

$$I_i^{(1)} = \frac{I_{\infty,i} e^{-\chi t}}{\chi (1 + V_i e^{-\chi t})^2} \sum_j \omega_{j \leftarrow i} (V_i - V_j) \times \left[ \chi t - \frac{V_i - V_j}{V_j} \ln \left( \frac{1 + V_j e^{-\chi t}}{1 + V_j} \right) \right]. \quad (19)$$

Summing this with the zeroth-order solution we reap the first-order perturbative approximation to the number of infected in city  $i$  at time  $t$ :

largest city, Auckland (*Tāmaki Makaurau*),  $N_{\text{Auck}}$ , is affected/or affects the remainder of NZ,  $N_{\text{rem}}$ , we recombine NZ into these two subpopulations (thereby assuming that each of these populations are large - whose explicit populations are stated in Fig. 2 - and well mixed). Realistic transport rates are  $\omega_{\text{rem} \leftarrow \text{Auck}} = 6.0 \times 10^{-3} \text{ day}^{-1}$ , and  $\omega_{\text{Auck} \leftarrow \text{rem}} = 3.1 \times 10^{-3} \text{ day}^{-1}$  [24]. We set initially the infected to  $I_{0,\text{Auck}} = 0.01 \times N_{\text{Auck}}$ , and  $I_{0,\text{rem}} = 0.026 \times N_{\text{rem}}$ , for Auckland and the remainder of NZ, respectively. With these parameters we note that validity criterion, Eq. (21), for Auckland, is  $C_{\text{rem} \leftarrow \text{Auck}} = 1.45$  and for the remainder of NZ is  $C_{\text{Auck} \leftarrow \text{rem}} = 0.74$ .

We solved the dynamics of the individual infected populations of this system numerically (by the Runge-Kutta 4<sup>th</sup> order algorithm), given the parameters above and compared the numerical results to our result in Eq. (20). The results are plotted in Fig. 2, and are compared with the scenario that all airports are shut down, which is also a measure of how effective the the zeroth-order solution, Eq. (8), is at estimating the time-evolution of the epidemic. We note that our analytical result conforms well to the numerical result, where the absolute residuals ( $= |I_{\text{numer.}}(t) - I_i(t)|$ , where the former is the numerical result, and the latter our result from Eq. (20)) between both are given in the inset. The zeroth-order, “no travel,” solution performs poorly at estimating the interim dynamics of the epidemic, especially underestimating the fraction of infected in Auckland (at its peak residual, this result is off by approximately  $0.03 \times N_{\text{Auck}} = 44580$  people.).

To assess the scope of our perturbation result in this system, we explore the phase space of the fraction of initially infected for each subpopulation - Fig. 3 - with respect to Auckland (the corresponding phase-space plot for the remainder of NZ is qualitatively comparable.). We linearly vary the fraction of infected from  $10^{-4} \rightarrow$

$2 \times 10^{-1}$ , and for each pair we calculate the perturbation approximation, and the numerical solution. The absolute residuals (see Fig. 2, inset) between the numerical and perturbation solutions are calculated; the area under which is found. This area is divided by the area under the numerical solution to give the absolute residue area ratio. These ratios are plotted as a “heat map” in Fig. 3, where the lighter the shade, the closer our analytical result is to the numerical solution. When the fraction of initially infected for each subpopulation, compared to the other, is near equal the absolute residue area ratio is relatively small, and therefore a good approximation. This is predicted by our quantitative benchmark, the validity criterion, whose contour lines have been overlaid upon the heat map. The usefulness of the first-order approximation lies in the off-diagonal circumstances, and for one such instance is shown, via a white circle, on Fig.

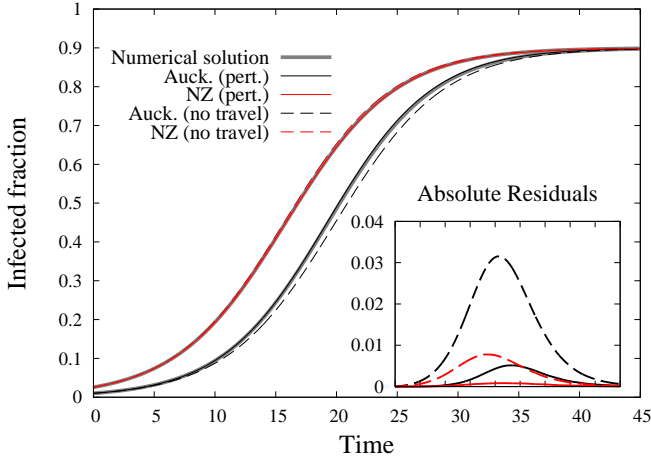


FIG. 2. The infected fraction of the populace over time (due to an SIS epidemic) for each city for an internationally quarantined New Zealand, apportioned into two subpopulations: Auckland, and the remainder of the country. Individuals (both susceptible and infected) may migrate between these two subpopulations only via commercial air transport; and through documented air transport data [25], we find the transport rates are effectively small - viz. Eq. (21). One clearly notes that in both populations, the first-order perturbation, Eq. (20) (solid red/black line), approximates the numerical solution well (solid gray line), compared to the zeroth-order solution, Eq. (8) (dashed red/black line), the case where infected migration is not considered. Transport rates (see main text) are:  $\omega_{\text{rem} \leftarrow \text{Auck}} = 6 \times 10^{-3} \text{ day}^{-1}$ , and  $\omega_{\text{Auck} \leftarrow \text{rem}} = 3 \times 10^{-3} \text{ day}^{-1}$  [24]. Validity criteria, Eq. (21), for each subpopulation are:  $C_{\text{rem} \leftarrow \text{Auck}} = 1.45$  and  $C_{\text{Auck} \leftarrow \text{rem}} = 0.74$  (see also Fig. 3). INSET: Absolute residuals of the first-order perturbation and the zeroth-order solution, with respect to the numerical solution (defined explicitly in the main text). Auxiliary parameters: population sizes (for the same year as the air traffic data),  $N_{\text{Auck}} = 1486000$ , and  $N_{\text{rem}} = 2919300$  [26]; virus parameters are:  $\beta = 0.25 \text{ day}^{-1}$ ,  $\gamma = 0.025 \text{ day}^{-1}$ . Initially infected populations for the two cities are:  $I_{0,\text{Auck}} = 0.01 \times N_{\text{Auck}}$ , and  $I_{0,\text{rem}} = 0.026 \times N_{\text{rem}}$ .

3, which is the pair of the initial fractions used for Fig. 2. These initial conditions demonstrate the outer limits of the approximation, where the zeroth-order has effectively failed, and the first-order is able still to approximate the numerical solution.

If we instead push the approximation past the recommended parameter sets, we can observe its breakdown. We present this in Fig. 4, using the initially infected values:  $(I_{0,\text{Auck}}, I_{0,\text{rem}}) = (0.0025 \times N_{\text{Auck}}, 0.05 \times N_{\text{rem}})$ . This point is shown on the phase-space plot, Fig. 3, as a white square.

## DISCUSSION AND CONCLUSION

Within this manuscript, we have derived a closed form, recursive, perturbative solution to an SIS epidemic on an

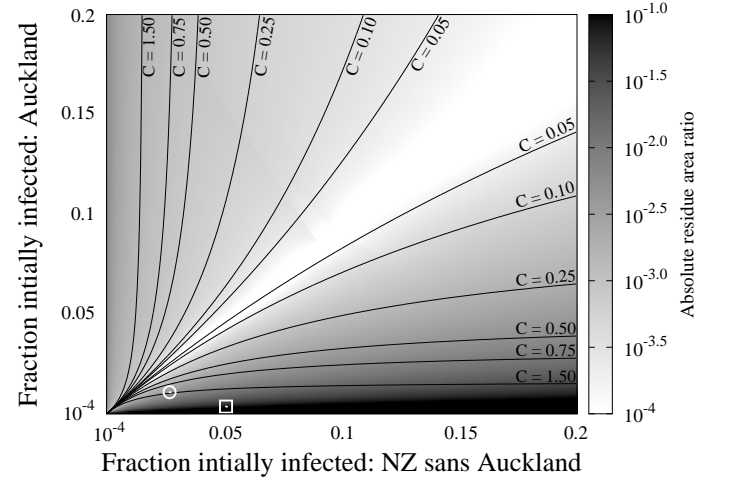


FIG. 3. Exploration of the phase space for the fraction of initially infected in the quarantined NZ system, with respect to Auckland. The lighter the shade, the closer the perturbative solution for the infected population in Auckland is to the numerical solution (explained below). Linear alteration of the fraction of initially infected for the Auckland population is shown on the ordinate, and for the remainder of NZ population on the abscissa (all other system parameters are kept constant - see Fig. 2). For each set of initial populations, we compare our perturbation result, Eq. (20), to the numerical solution. The absolute residues for Auckland (see Fig. 2 inset) between these two results are calculated, then the area under which is measured and divided by the integration of the numerical solution for the same time-frame. The absolute residue area ratio is plotted as a heat map, where the shade (see color bar above) denotes the magnitude of the value. The initial conditions used for Fig. 2 are marked with a white circle in the lower left quadrant. An instance when the first-order perturbation breaks down is highlight via a white square, and whose explicit temporal evolution is shown in Fig. 4. Superimposed on the initial condition heat map is the corresponding contour plot of the validity criterion, Eq. (21) (where  $C$  denotes  $C_{\text{rem} \leftarrow \text{Auck}}$ ). The phase space heat map for the remainder of NZ is qualitatively similar.

arbitrary network, stated in Eqs. (13) and (15). We have proceeded to explicitly calculate the first-order perturbation to the population of infected persons in the  $i^{\text{th}}$  city as a function of time, to wit, Eq. (20), and then provided a quantitative benchmark under what conditions this solution is accurate: the validity criterion, Eq. (21).

To verify our derived results we simulated an SIS epidemic on an internationally quarantined New Zealand. This comparison served a two-fold objective; firstly the use of documented air traffic data [25] showed that in this medium of transport, our base assumption: transport rates between communities are small, is indeed reasonable and accurate as a perturbation parameter (see main text for further discussion, caveats, and explicit numbers). Secondly, when the system parameters give a validity criterion near or less than unity, the first-order solution, Eq. (20), approximates the numerical solution well, as shown Fig. 2. The range of usefulness of the validity criterion and its dependence on the fraction of initially infected is explored in Fig. 3, and a set of parameters for when the approximation breaks down is provided in Fig. 4.

The derivation and the subsequent verification through the numerical simulations assumed detailed balance and has leaned on the assumption that a finite fraction of each city is initially infected; the derivation then seeks to describe temporal evolution of the epidemic. In this case, this would apply to the reasonable scenario where the authorities of a region would look to understand the dynamics of the epidemic once their community of cities

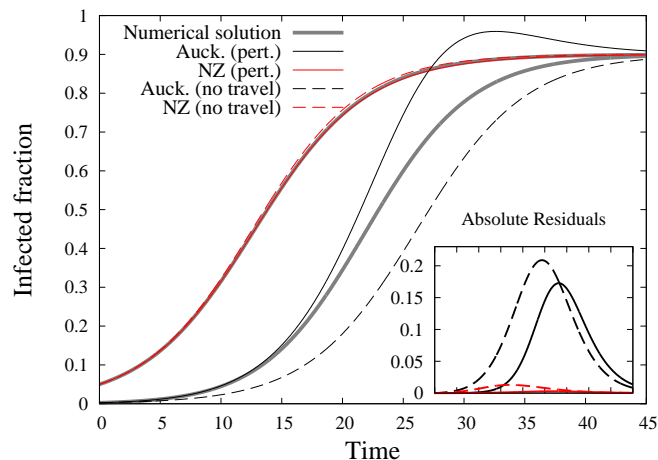


FIG. 4. Demonstration of the breakdown of Eq. (20), our first-order perturbation, with initial infection parameters corresponding to a large validity criterion. In this system the initially infected for Auckland and the remainder of NZ are:  $I_{0,\text{Auck}} = 0.0025 \times N_{\text{Auck}}$  and  $I_{0,\text{rem}} = 0.05 \times N_{\text{rem}}$  respectively (this pair is marked as a white square in the phase-space for Fig. 3). The corresponding validity criteria are:  $C_{\text{rem} \leftarrow \text{Auck}} = 9.14$ ,  $C_{\text{Auck} \leftarrow \text{rem}} = 4.65$ . For the remaining system parameters and explanation of inset see Fig. 2 caption.

is found to be infected. This of course shifts the problem from the usual discourse found within the literature, where the infection is present initially in one population and spreads, such as [27]. It would be of interest to understand how higher order perturbation terms may regularize our first-order approximation.

One of the striking benefits of the solution is that one has an analytical expression for  $I_i(t)$  at all times and as such naturally out-performs usual investigative methods of numerical integration. In this way this solution can be used to gauge parameter sets of large numerical simulations.

The derivation makes no assumptions on the type of the network, whether it be a real-world network, regular lattice, a random E-R graph, or a scale-free network [9]. It has also only assumed that the population of each node is large enough such that the mean-field nature of Eq. (3) is true. Therefore the nodes may be seen as communities or countries, rather than only cities. This generality is advantageous for future investigations of SIS epidemics on complex networks/metapopulations as this work may be used in parallel as a confidence measure.

These calculations were built upon the assumption of large populations, where mean-field approximations are valid. A natural extension would be the effect of stochasticity for low populations. This may be found through an analytic perturbative solution of the associated master equation akin to that defined for an SIR epidemic in the work of Hufnagel et al. [27].

Although the analytics have been developed under the guise of epidemic modeling, this mathematical framework may be conveniently adopted by other interdisciplinary fields with population growth and metapopulation structure, for example Theoretical Ecology and the concept phenomenon of island colonization [28].

In conclusion, we hope the mathematical framework determined herein will shift part of the academic interest of epidemics on networks from large scale numerical simulations back to the bedrock of analytical analysis.

The authors would like to thank Bo Söderberg, Erik Lagerstedt, Olivia Woolley-Meza and Sigurður Æ. Jónsson for stimulating discussions regarding the work herein.

\* Corresponding author: lloyd.sanders@thep.lu.se; Department of Astronomy and Theoretical Physics, Lund University, SE-223 62 Lund, Sweden

† Northwestern Institute on Complex Systems and Department of Engineering Sciences and Applied Mathematics, Northwestern University, Evanston, IL, USA

‡ Department of Astronomy and Theoretical Physics, Lund University, SE-223 62 Lund, Sweden

[1] W. O. Kermack and A. G. McKendrick, Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences, **115**, 700 (1927).

- [2] H. W. Hethcote, in *Three basic epidemiological models*, Applied Mathematical Ecology, Vol. 18 (Springer Berlin Heidelberg, 1989) pp. 119–142.
- [3] K. M. E. Turner, E. J. Adams, N. Gay, A. C. Ghani, C. Mercer, and W. J. Edmunds, Theoretical biology & medical modelling, **3**, 3 (2006).
- [4] J. Llibre and C. Valls, Journal of Mathematical Analysis and Applications, **344**, 574 (2008).
- [5] P. Das, Z. Mukandavire, C. Chiyaka, A. Sen, and D. Mukherjee, Differential Equations and Dynamical Systems, **17**, 393 (2010).
- [6] V. Méndez, D. Campos, and W. Horsthemke, Physical Review E, **86** (2012).
- [7] D. T. Gillespie, Annual review of physical chemistry, **58**, 35 (2007).
- [8] M. J. Keeling and J. V. Ross, Journal of the Royal Society, Interface / the Royal Society, **5**, 171 (2008).
- [9] M. E. J. Newman, SIAM Review, **45**, 167 (2003).
- [10] O. Woolley-Meza, C. Thiemann, D. Grady, J. Lee, H. Seebens, B. Blasius, and D. Brockmann, The European Physical Journal B - Condensed Matter and Complex Systems, **1** (2011).
- [11] J. P. Bagrow and Y.-R. Lin, PloS one, **7**, e37676 (2012).
- [12] D. Brockmann, in *Reviews of Nonlinear Dynamics and Complexity*, edited by H. G. Schuster (Wiley-VCH, 2009) pp. 1–24, ISBN 9783527408504.
- [13] I. Hanski, Nature, **396**, 41 (1998).
- [14] B. Grenfell and J. Harwood, Trends in ecology & evolution, **12** (1997).
- [15] H. Lund, L. Lizana, and I. Simonsen, Journal of Statistical Physics, **151**, 367 (2013).
- [16] V. Colizza and A. Vespignani, Physical Review Letters, **99**, 1 (2007).
- [17] R. Parshani, S. Carmi, and S. Havlin, Physical Review Letters, **104**, 258701 (2010).
- [18] P. Van Mieghem and R. van de Bovenkamp, Physical Review Letters, **110**, 108701 (2013).
- [19] S. C. Ferreira, C. Castellano, and R. Pastor-Satorras, Physical Review E, **86**, 041125 (2012).
- [20] F. Arrigoni and A. Pugliese, Journal of mathematical biology, **45**, 419 (2002).
- [21] F. Ball, Mathematical Biosciences, **156**, 41 (1999).
- [22] M. Abramowitz and I. A. Stegun, *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables*, 9th ed. (Dover, 1972).
- [23] It should be noted that  $\sum_j C_{j \leftarrow i}$  may be within the valid range for city  $i$ , but the corresponding ratio sum for some other city  $j$  may not be. In this case, Eq. (20) would be reasonable still for  $I_i(t)$ , but not for  $I_j(t)$ , therefore caution is advised. This is apparent upon the inspection of the absolute residues inset on either figure. One notes the difference for one city is higher than that for the other.
- [24] The air transport rates are calculated from the worldwide air traffic data provided by Ref. [25], taken the same year as the Subnational census data of NZ [26]. We have used the average number of *available* seats (from the total average flights) per day as the measure of the number of people commuting between two airport nodes in the NZ network. We have also removed all international connections to and from NZ airports.
- [25] OAG, “Air Traffic Data,” <http://www.oag.com> (2011), [Data retrived: 03-Feb-2011].
- [26] Statistics New Zealand, “Subnational Populations Estimates: At 30 June 2011,” [http://www.stats.govt.nz/browse\\_for\\_stats/population/estimates](http://www.stats.govt.nz/browse_for_stats/population/estimates) (2011), [Online; accessed 4-April-2013].
- [27] L. Hufnagel, D. Brockmann, and T. Geisel, Proceedings of the National Academy of Sciences of the United States of America, **101**, 15124 (2004).
- [28] R. Levins, Bulletin of the Entomological Society of America, **15**, 237 (1969).